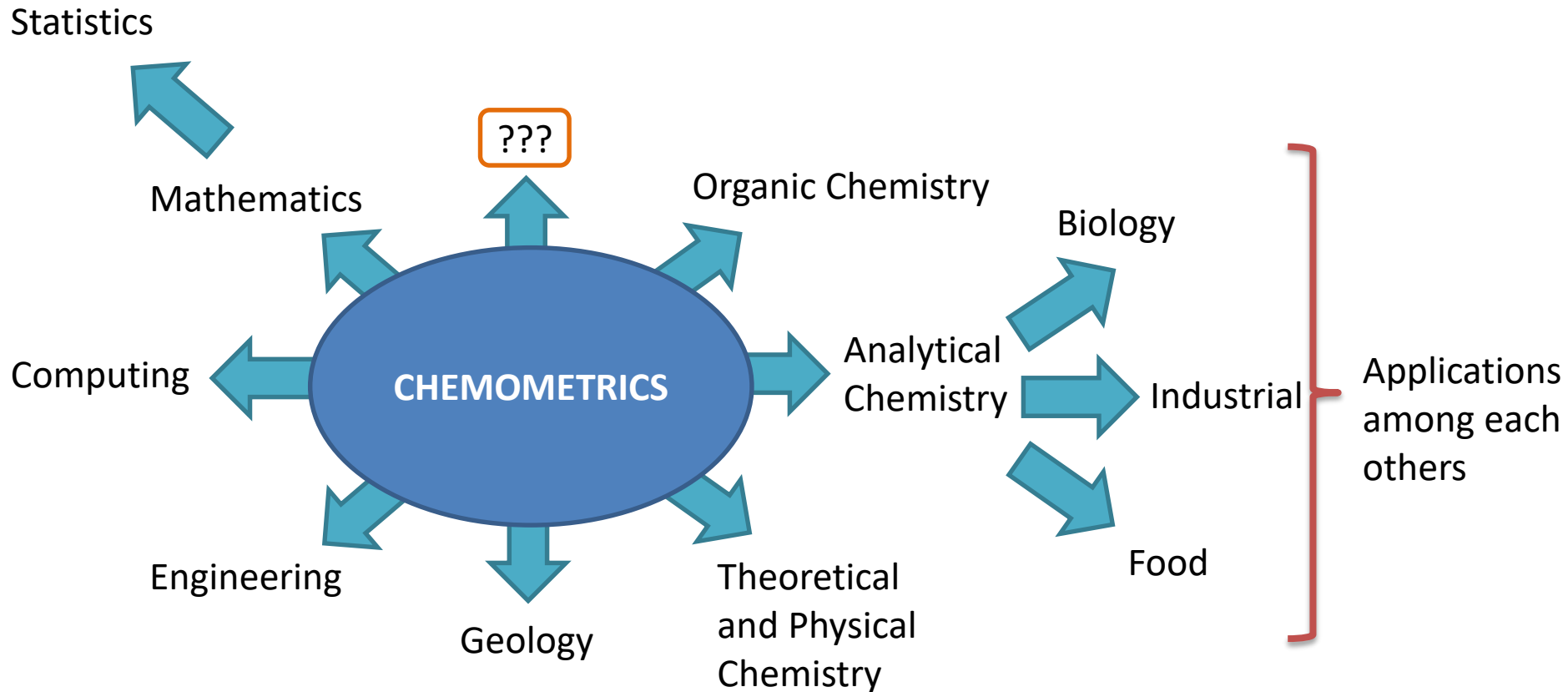


Introduction to chemometrics



ศูนย์วิจัยเทคโนโลยีหลังการเก็บเกี่ยว

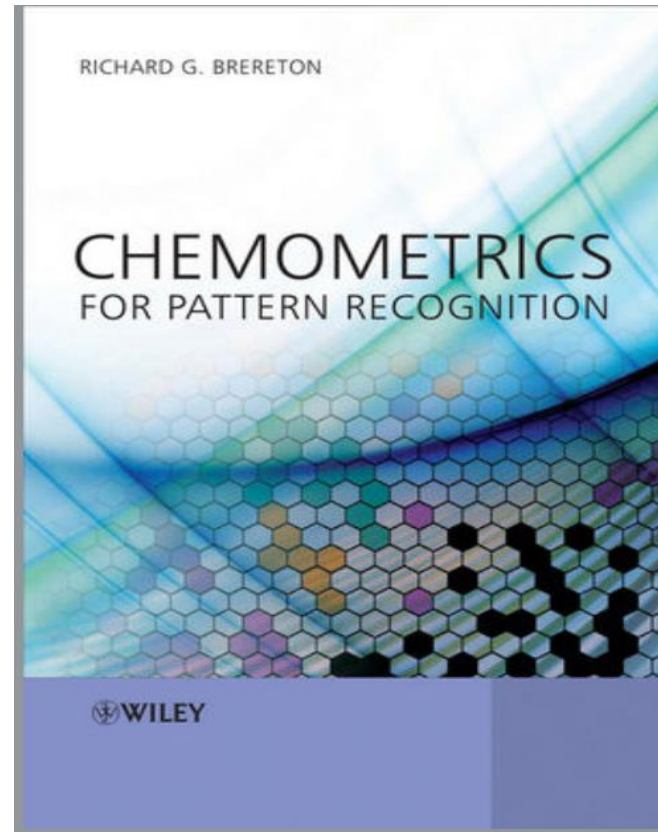
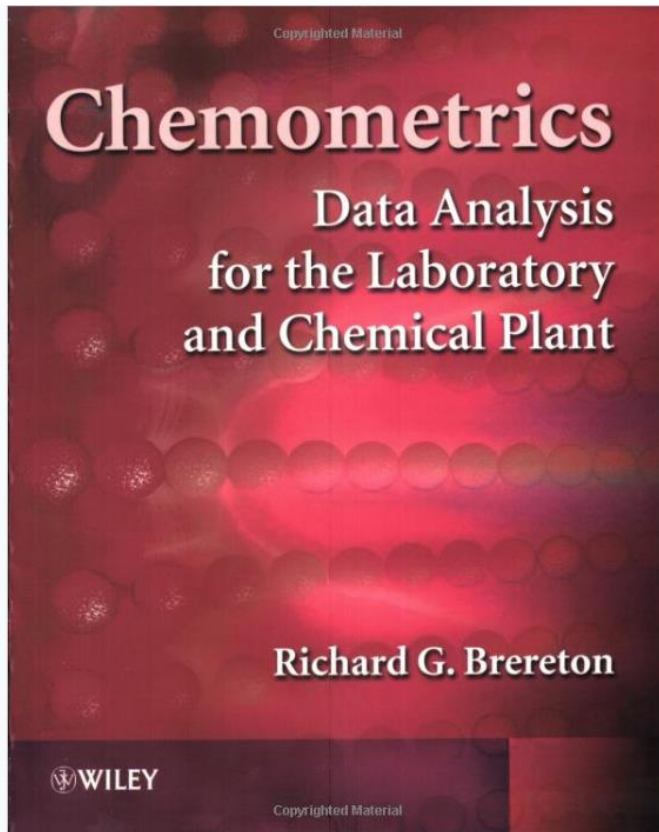
คณะเกษตรศาสตร์ มหาวิทยาลัยเชียงใหม่ (31 พ.ค. -1 มิ.ย. 2561)

ผศ.ดร. ศิลา กิตติวัชนะ และคณะนักศึกษาคณาจารย์

ภาควิชาเคมี คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

E-mail: silacmu@gmail.com

Tel: 087-9166692



Some important references

Outline

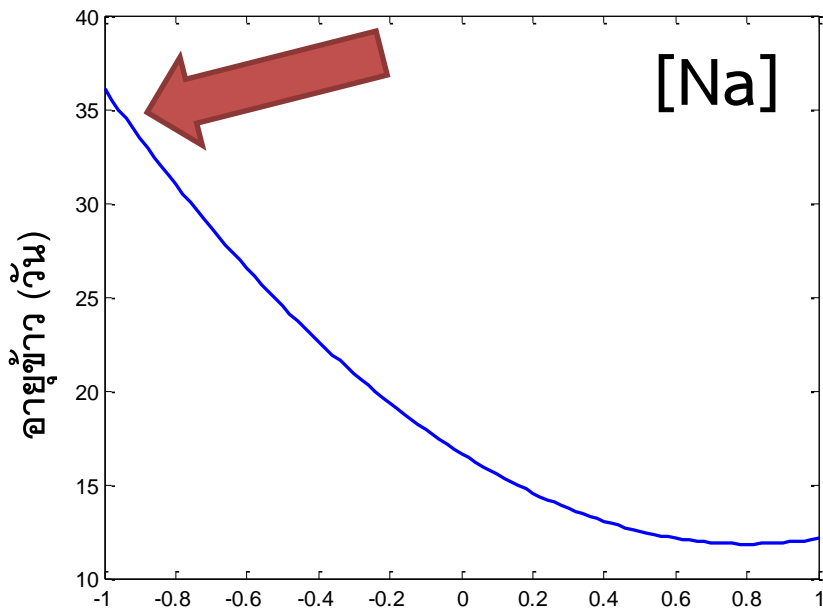
- Introduction to **chemometrics**
- Data analysis using **PCA**
- **Data preprocessing**
- **NIR** analysis for agricultural products
- Calibration modeling using **MLR, PCR** and **PLS**

Univariate data analysis

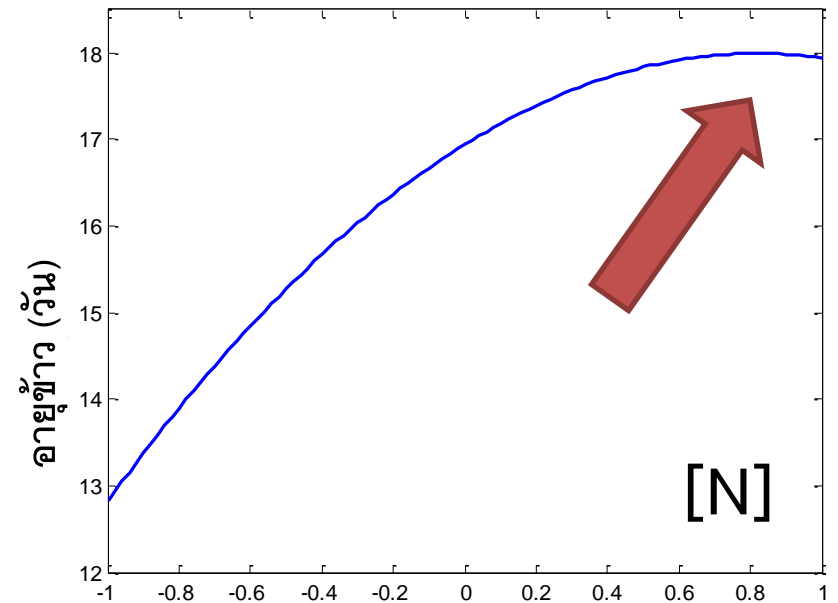
VS

Multivariate data analysis

การศึกษาผลของปริมาณ N และ Na ที่มีต่อ คุณภาพความหอมของเมล็ดข้าวหอมมะลิ



ความเข้มข้นของเกลือโซเดียม (Na)

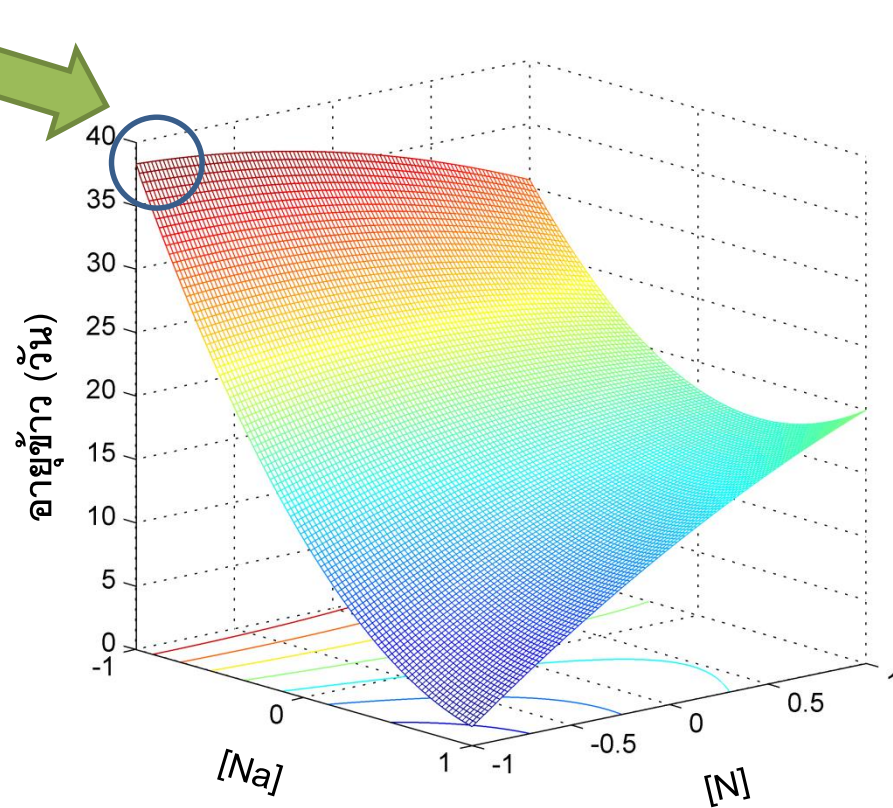
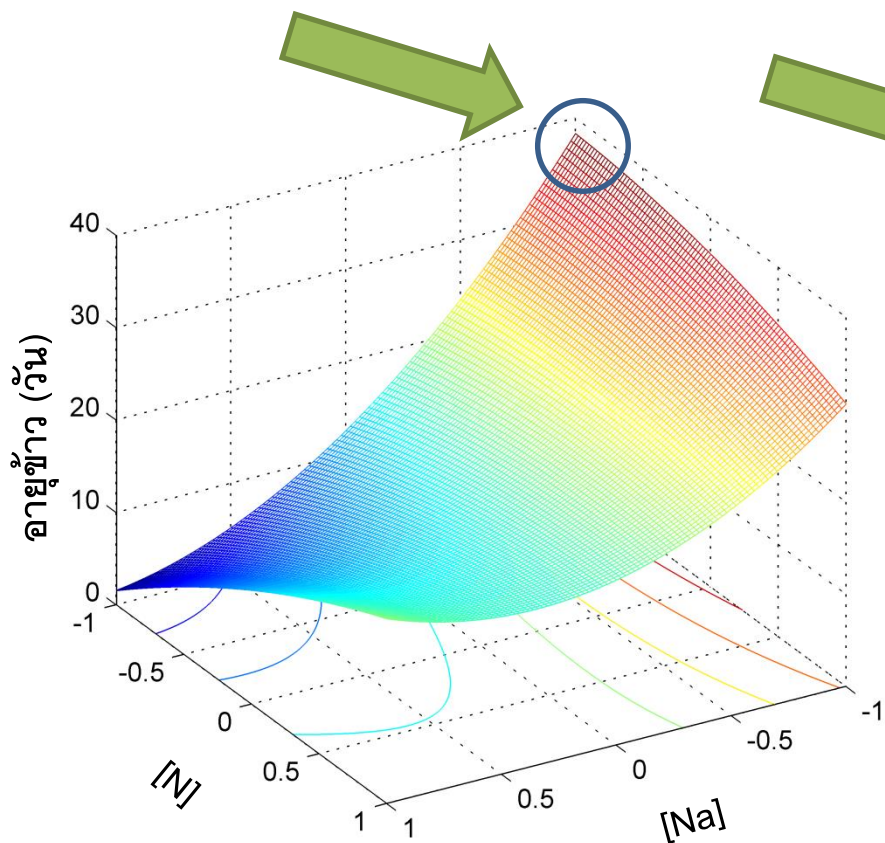
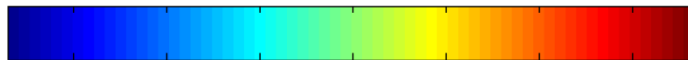


ความเข้มข้นปุ๋ยไนโตรเจน (N)

.....ข้าวจะมีอายุยืนที่สุดถ้า Na มีความเข้มข้นต่ำที่สุด (-1) แต่ N จะต้องมีความเข้มข้นสูงๆ (+1).....

Sooner death

Longer life



แต่จริงๆแล้ว.....ข้าวจะมีอายุยืนที่สุดถ้าทั้ง Na และ N มีความเข้มข้นต่ำที่สุด (-1)

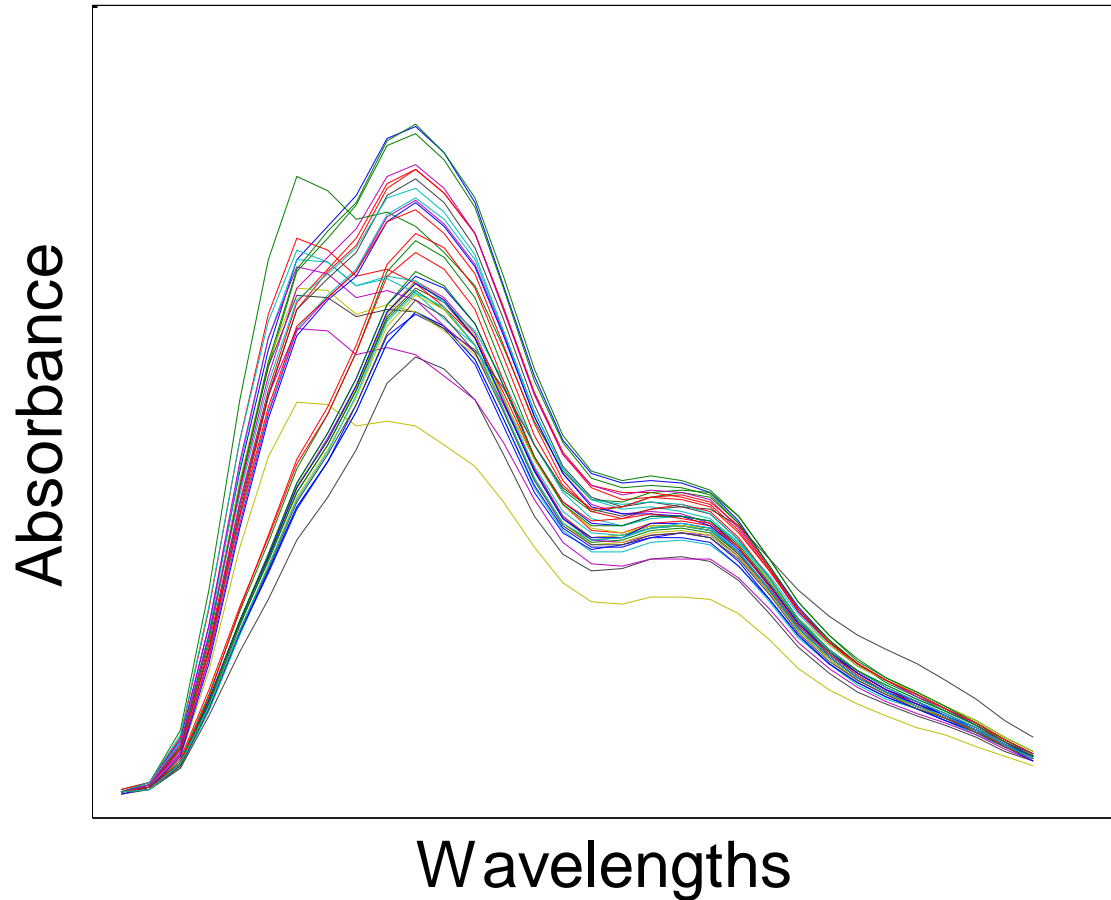
Univariate data analysis

- One-variable-at-a-time analysis
- It is simple but may not correctly reflect the exact situation.

Multivariate data analysis

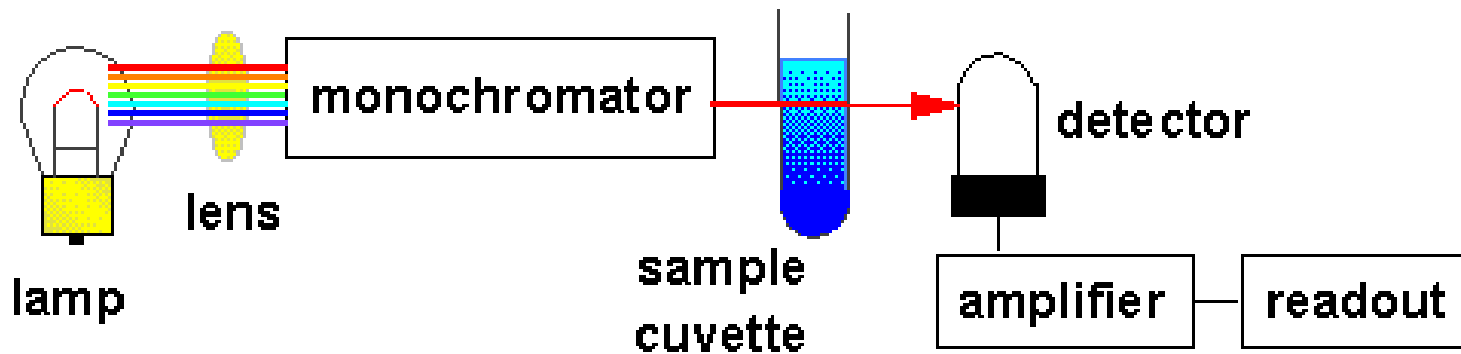
- (All of) the variables are simultaneously used in the analysis.
- It is more complicated but we probably get what we actually want.
- So, we do need a good strategy and it is chemometrics

A case study

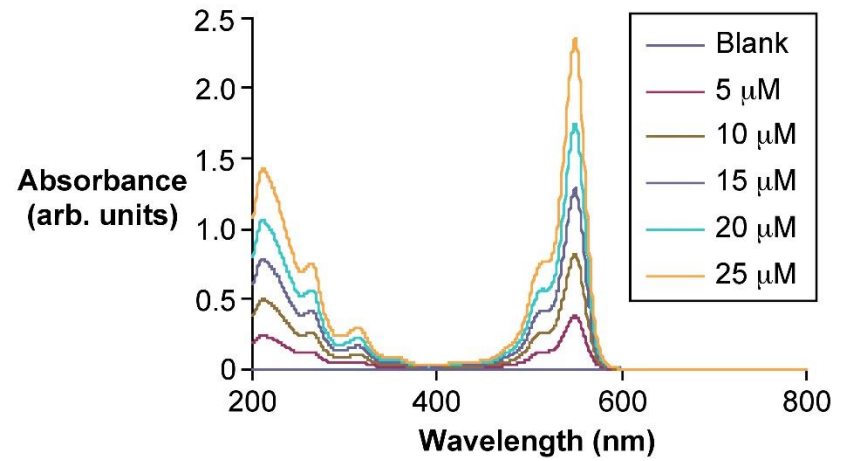


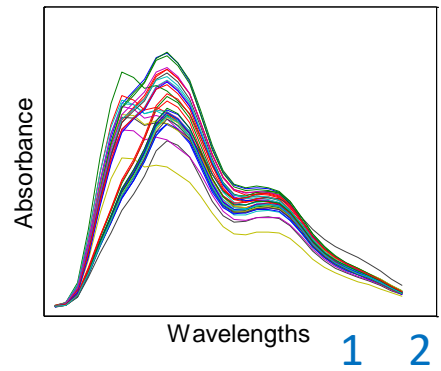
- **UV/Vis spectra of 39 samples.**
- **The absorbance was measured at 24 different wavelengths from 300-800 nm.**

UV-Vis spectrometer



©1995 CHP





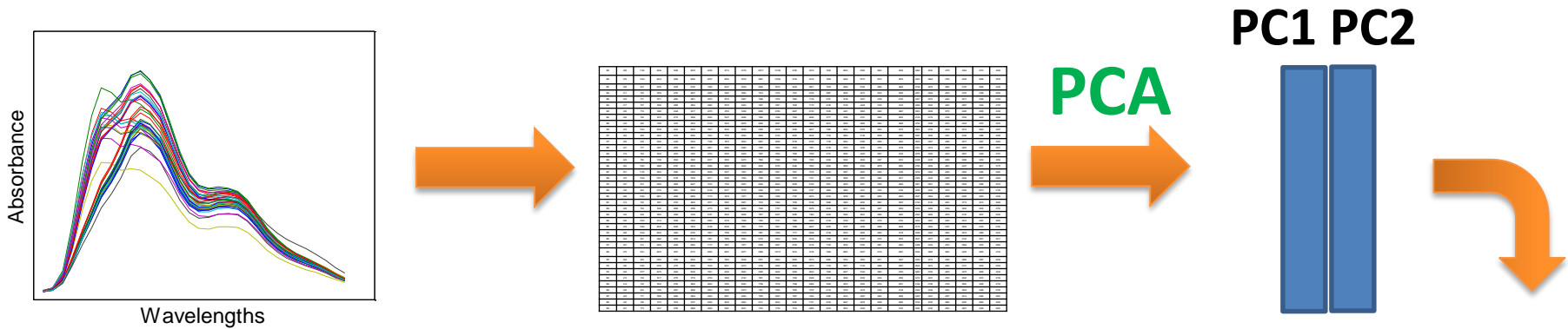
$x_{1,24}$ is the absorbance of the 1st sample at the 24th wavelength

1
2
3
4
.....
Samples
.....
36
37
38
39

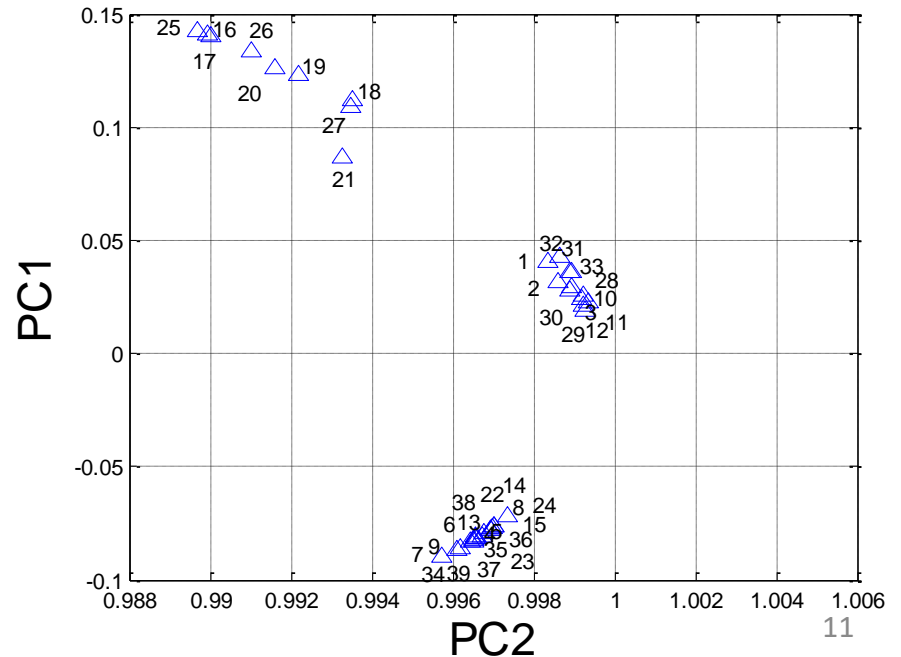
Wavelengths

	1	2	3	4																21	22	23	24
1	385	498	1128	2826	5098	6878	8235	8714	9179	10017	10188	9792	9074	7832	6500	5565	5063	4938	4962	4930	4779	4364	3710	3099
2	385	476	1033	2633	4869	6656	8037	8552	9018	9897	10099	9704	9016	7809	6443	5512	5010	4876	4909	4864	4742	4328	3696	3078
3	381	458	947	2376	4478	6180	7495	8047	8433	9292	9573	9224	8601	7543	6266	5349	4856	4699	4736	4679	4569	4197	3628	3036
4	353	411	723	1553	2700	3637	4585	5262	5982	7013	7472	7241	6743	5867	4876	4212	3919	3909	4061	4084	4009	3700	3189	2659
5	356	424	771	1671	2880	3851	4873	5579	6367	7468	7916	7684	7143	6181	5138	4447	4138	4155	4307	4315	4244	3911	3339	2793
6	354	417	757	1656	2858	3850	4880	5578	6350	7427	7881	7649	7119	6196	5138	4448	4133	4142	4269	4294	4225	3887	3344	2787
7	348	409	716	1480	2444	3217	4075	4703	5438	6386	6786	6607	6150	5335	4441	3863	3622	3667	3800	3839	3772	3485	2995	2498
8	350	428	792	1672	2712	3599	4550	5237	6086	7092	7426	7224	6696	5771	4786	4191	3940	4002	4133	4179	4130	3769	3223	2689
9	356	434	826	1799	3019	4053	5133	5933	6876	8060	8497	8267	7666	6612	5508	4808	4520	4568	4720	4786	4703	4324	3707	3076
10	378	472	1022	2548	4632	6304	7607	8093	8539	9359	9578	9226	8607	7490	6252	5375	4903	4775	4802	4728	4616	4219	3624	3027
11	377	464	988	2422	4414	5982	7204	7700	8092	8893	9143	8822	8232	7192	5984	5149	4676	4560	4587	4526	4412	4046	3485	2908
12	377	459	936	2330	4333	5934	7166	7674	8054	8847	9098	8745	8159	7135	5924	5056	4599	4462	4500	4457	4336	3971	3427	2864
13	355	421	777	1687	2878	3841	4851	5568	6366	7452	7904	7643	7081	6145	5105	4435	4125	4149	4304	4314	4239	3905	3357	2801
14	354	419	763	1640	2814	3790	4794	5505	6308	7387	7822	7595	7059	6131	5095	4403	4097	4122	4276	4293	4223	3891	3343	2781
15	354	423	785	1706	2904	3890	4917	5653	6473	7569	7996	7809	7295	6363	5332	4639	4327	4312	4438	4445	4356	3989	3433	2856
16	396	522	1290	3368	6167	8223	9456	9262	8830	8933	8725	8346	7852	6932	5858	5093	4679	4650	4789	4839	4803	4427	3801	3175
17	387	501	1176	3053	5565	7430	8545	8377	7997	8081	7881	7521	7075	6264	5287	4600	4230	4186	4316	4362	4308	3992	3425	2862
18	384	472	1034	2721	5158	7059	8211	8199	7832	7978	7905	7579	7191	6451	5498	4791	4392	4311	4408	4445	4376	4068	3547	2984
19	375	457	972	2459	4609	6227	7214	7192	6837	6932	6824	6509	6160	5521	4679	4051	3732	3693	3801	3821	3796	3528	3066	2573
20	364	435	894	2172	3981	5305	6139	6077	5779	5849	5757	5493	5187	4660	3968	3441	3163	3136	3242	3240	3218	2998	2610	2199
21	380	469	1044	2661	4929	6659	7713	7674	7371	7506	7458	7197	6910	6316	5476	4861	4540	4453	4547	4564	4538	4295	3823	3348
22	351	413	740	1581	2704	3625	4571	5250	5997	7015	7464	7259	6738	5886	4909	4252	3967	3975	4122	4140	4054	3719	3213	2686
23	354	428	793	1692	2845	3789	4773	5485	6288	7349	7757	7508	6979	6024	5011	4346	4067	4093	4243	4269	4186	3842	3303	2750
24	354	442	849	1863	3094	4136	5197	5934	6849	7977	8331	8085	7480	6433	5329	4641	4363	4400	4554	4606	4518	4140	3537	2940
25	384	506	1209	3113	5546	7324	8385	8178	7848	7930	7721	7385	6960	6143	5189	4520	4179	4152	4254	4281	4240	3919	3357	2809
26	380	479	1098	2832	5240	7053	8138	8030	7661	7772	7621	7259	6847	6076	5147	4439	4088	4046	4170	4202	4147	3845	3319	2780
27	379	468	1034	2654	4963	6730	7807	7758	7419	7551	7459	7171	6839	6149	5274	4605	4252	4178	4288	4302	4269	3974	3480	2942
28	380	464	991	2462	4552	6214	7481	7959	8350	9172	9404	9054	8416	7335	6078	5187	4710	4576	4641	4577	4468	4102	3521	2917
29	377	451	910	2254	4249	5862	7117	7637	7977	8788	9061	8720	8134	7105	5916	5055	4591	4463	4488	4433	4313	3964	3415	2860
30	384	475	1021	2603	4870	6712	8085	8657	9077	9969	10212	9814	9135	7969	6605	5637	5116	4981	5026	4968	4820	4429	3800	3159
31	375	464	994	2453	4456	6064	7250	7682	8036	8781	8964	8625	8013	6960	5790	4953	4511	4407	4459	4419	4329	3966	3397	2834
32	379	490	1095	2700	4770	6393	7618	8021	8419	9156	9295	8934	8313	7183	5971	5128	4698	4597	4643	4612	4490	4101	3509	2933
33	379	484	1077	2679	4824	6526	7821	8248	8680	9455	9642	9269	8623	7489	6227	5334	4880	4763	4820	4799	4676	4275	3659	3049
34	352	415	749	1617	2770	3719	4709	5409	6192	7261	7692	7492	6949	6032	4996	4335	4034	4059	4210	4241	4160	3840	3310	2746
35	354	414	742	1602	2760	3684	4654	5350	6090	7156	7618	7398	6868	5993	4974	4302	4016	4021	4166	4196	4135	3800	3283	2722
36	353	428	802	1726	2901	3856	4853	5558	6373	7441	7865	7625	7072	6134	5108	4418	4132	4139	4322	4336	4250	3904	3348	2792
37	357	424	777	1685	2892	3894	4920	5651	6465	7578	8050	7827	7284	6365	5312	4597	4294	4299	4425	4441	4357	4013	3462	2891
38	358	430	796	1774	3073	4157	5269	6063	6962	8141	8624	8397	7791	6784	5647	4878	4548	4569	4720	4765	4666	4290	3682	3062
39	353	421	773	1647	2765	3688	4669	5391	6231	7315	7733	7514	6965	6023	5003	4343	4075	4120	4280	4330	4254	3907	3349	2800

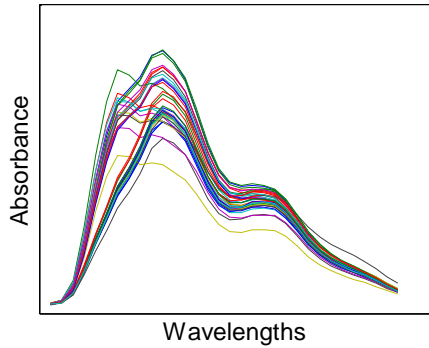
Principal Component Analysis (PCA)



Score plot of the 39 spectrum data (PC1 vs PC2)



Self Organising Map (SOM)

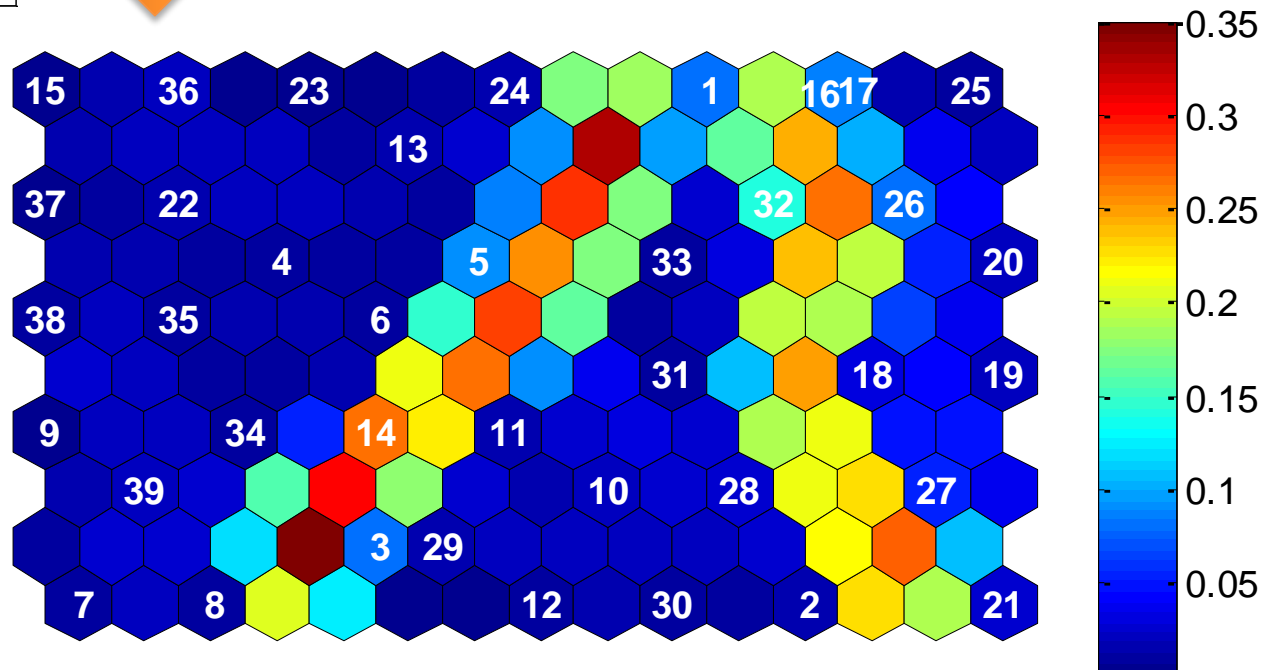


SOM

(one of the most conventional ANNs)



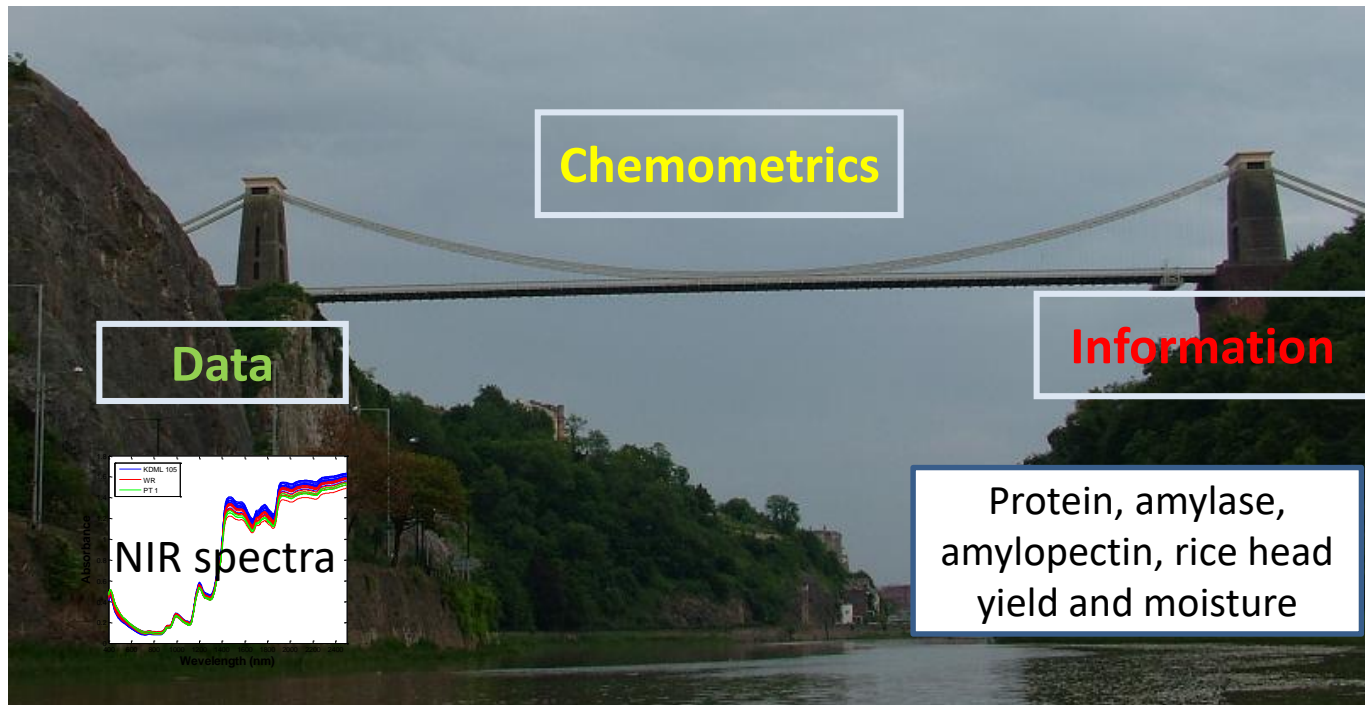
Dissimilarity



SOM map of the 39 spectrum data

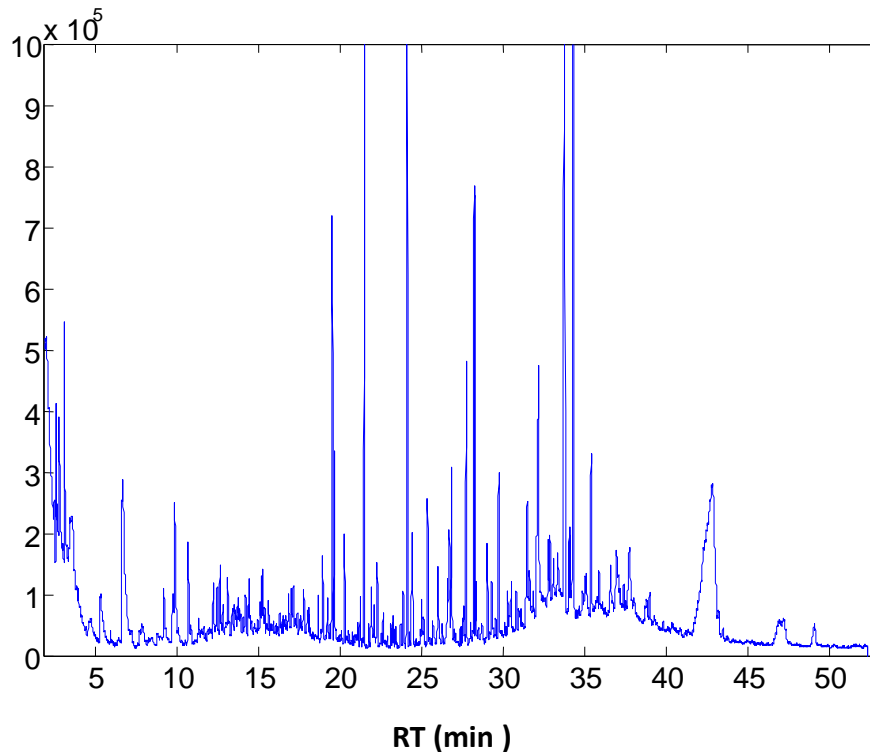
- *“Chemometrics has been defined as the application of mathematical and statistical methods to chemical measurements.”* **in 1980 by Kowalski.**
- *“Chemometrics is a chemical discipline that uses mathematics, statistics and formal logic*
 - (a) to design or select optimal experimental procedures;*
 - (b) to provide maximum relevant chemical information by analysing chemical data; and*
 - (c) to obtain knowledge about chemical systems”***in 1997 by Massart.**

Connecting analytical data to an useful information.....



Data	Information
UV/Vis spectra	<ul style="list-style-type: none"> • ความเข้มข้นของสีผสมอาหารในน้ำหวาน
Soil compositions (major and trace elements)	<ul style="list-style-type: none"> • ปริมาณสารหอมในเมล็ดข้าวหอมมะลิ • ดินใช้ปลูกอาโวคาโดได้ดีหรือไม่?
GC-MS spectra of blood samples	<ul style="list-style-type: none"> • ความถี่ในการสูบบุหรี่ของผู้ป่วย • ชนิดของโรคมะเร็งที่กำลังเป็น • ความสุขของมนุษย์
GC-MS spectra from apples	<ul style="list-style-type: none"> • อายุของแอปเปิ้ลที่วางขายในห้างสรรพสินค้า

Modern developments



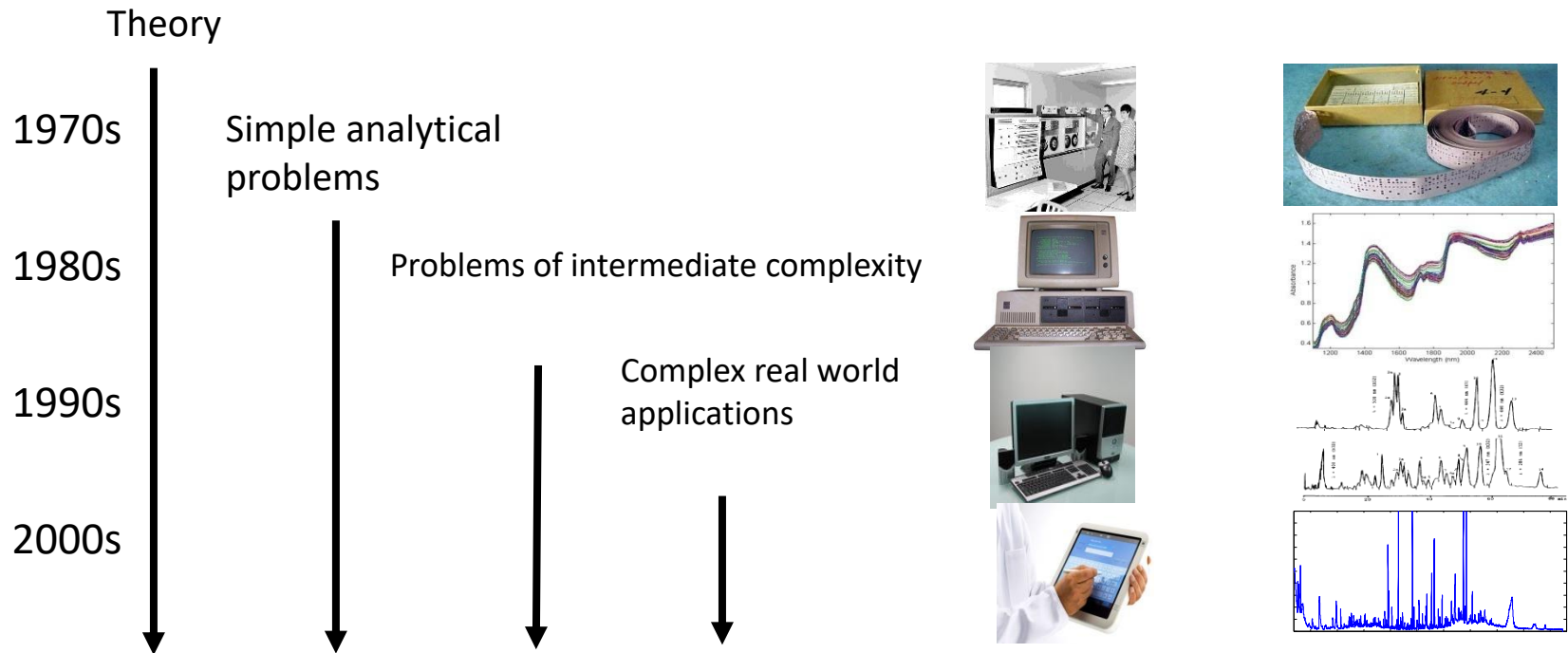
- **GCMS profiling in biology**
 - Human sweat
 - ✓ 500 + detectable compounds
 - ✓ Many origins
 - Microbes
 - Food eaten
 - Psychological state
 - Genetics
- **Very complex datasets and patterns**

Modern developments

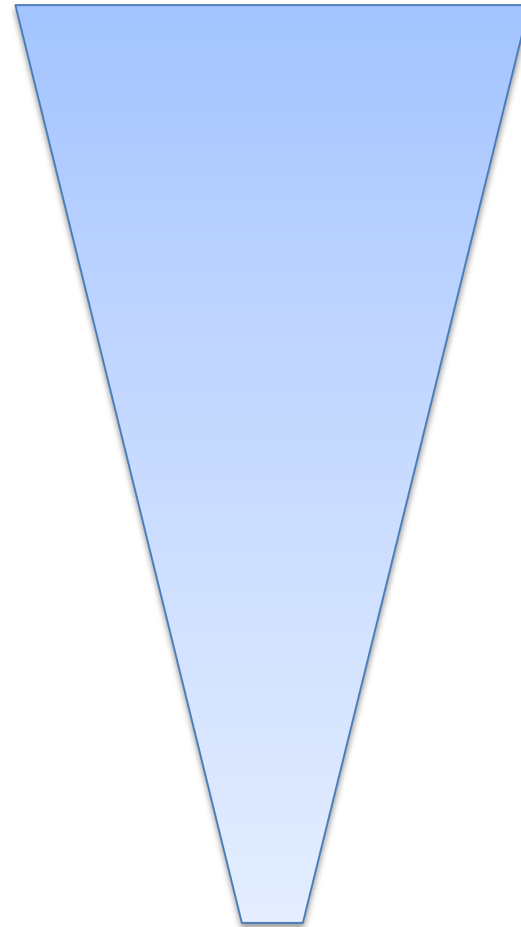
- **Possible to acquire huge datasets**
- **Example : 1,000 GCMSs of 500 compounds each 500,000 peaks**
 - 5 min to identify, check, integrate a peak = 41,666 hours work = 18.5 years at 45 weeks a year, 50 hours a week
- **You do need chemometrics**



Summary timeline



- Big Data
- Data mining
- Chemometrics
- Artificial intelligent
(ปัญญาประดิษฐ์, AI)
- Pattern recognition
- Machine learning
- Artificial neural networks
(ระบบเครือข่ายปัญญาเทียม, ANNs)

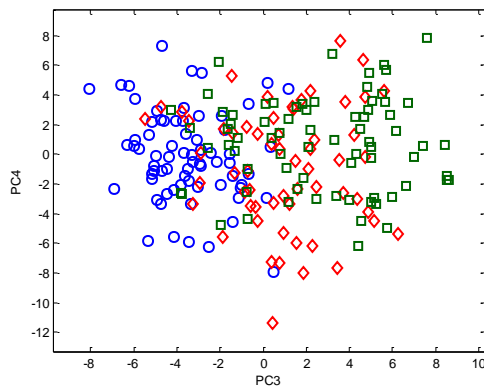


Applications of Chemometrics

- **Data exploratory**

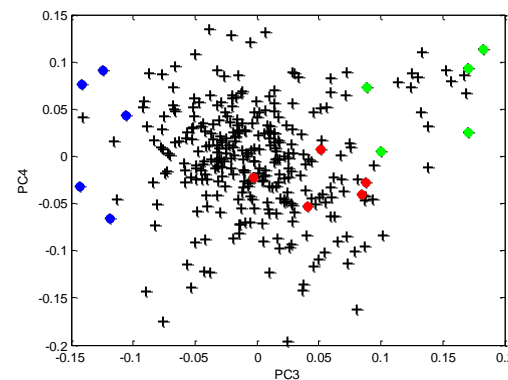
What does the data look like?

Mouse urine – genetics



Scores (mouse samples)

- Class A : AKRH2k
- ◇ Class B : B6H2k
- Class C : B6H2b



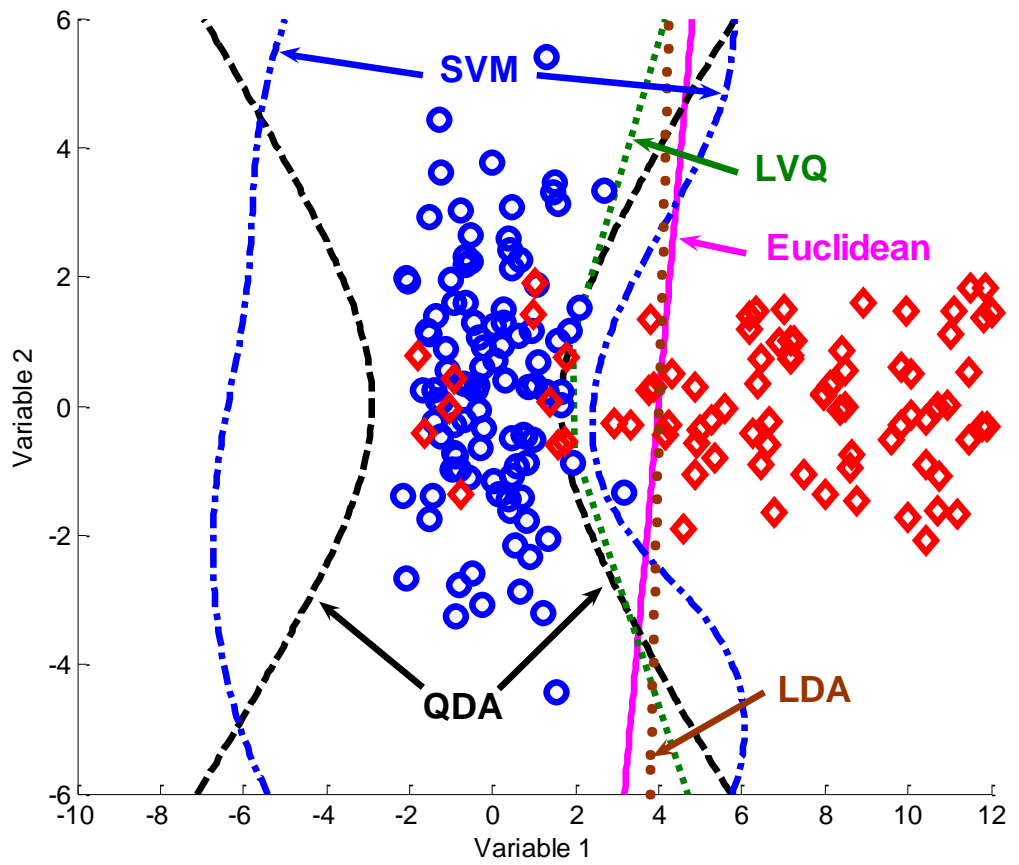
Loadings (chemicals detected)

Chemicals most diagnostic of each group marked in colour

• Classification/Pattern recognition

What is the sample category?

Is it possible to predict membership of an unknown sample?

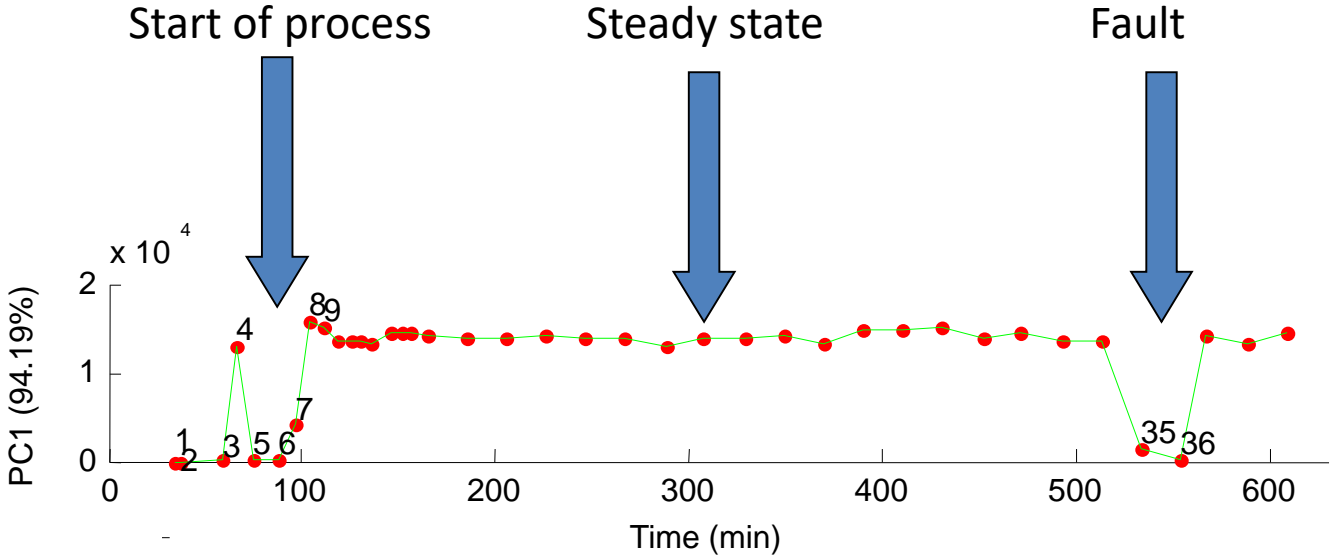


• Process monitoring/control

Is the process working OK?

When is the fault detected and what is the cause?

Is it possible to predict if the fault will be occurred...

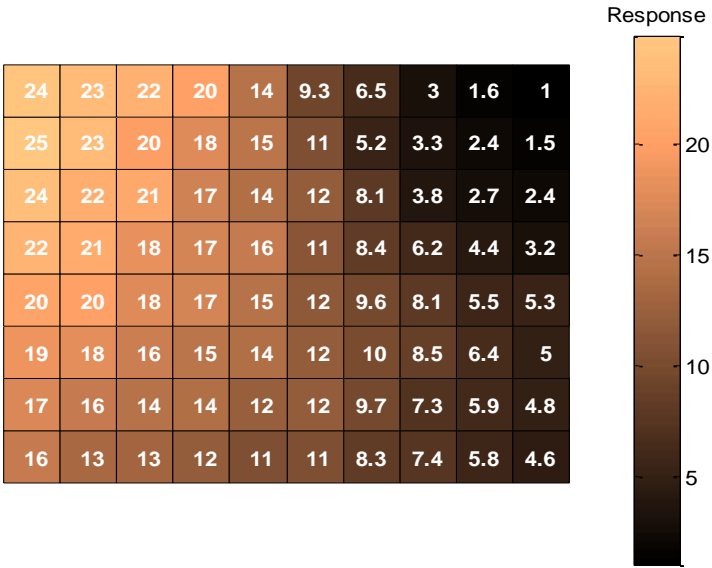
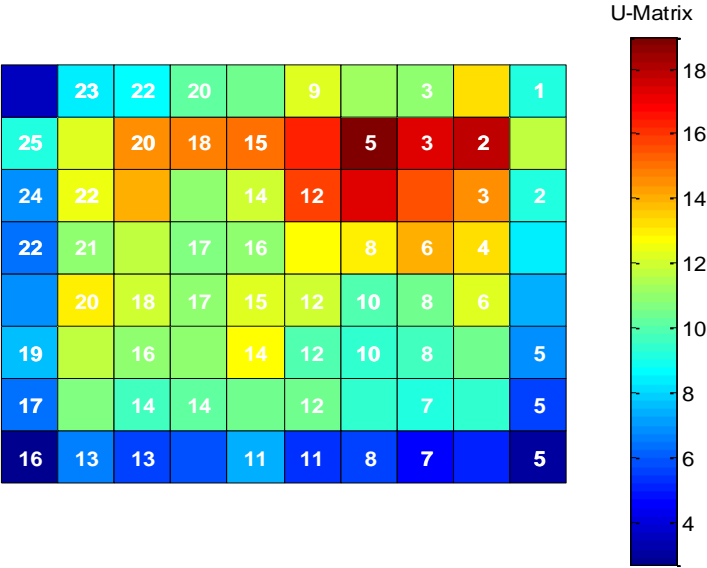


4

Multivariate calibration

What is sample's concentration?

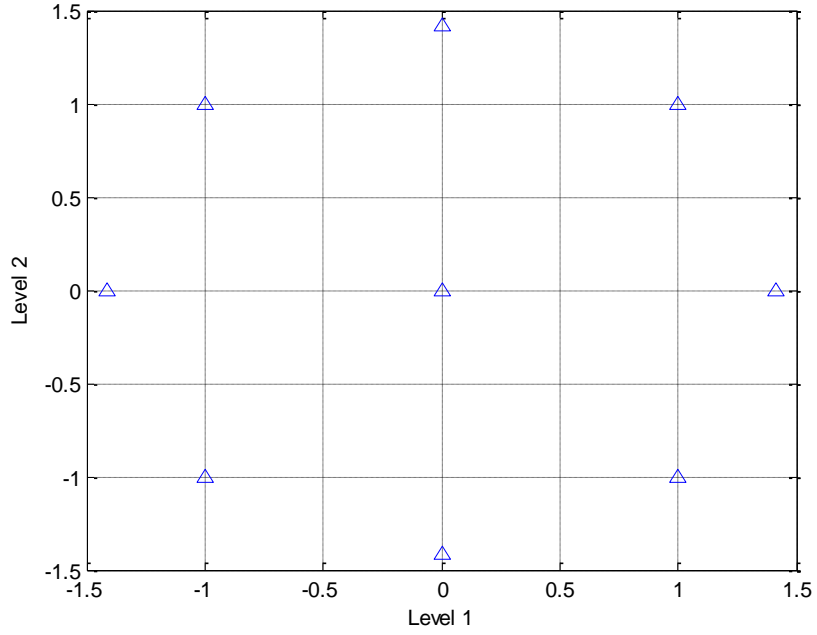
How old is this sample?

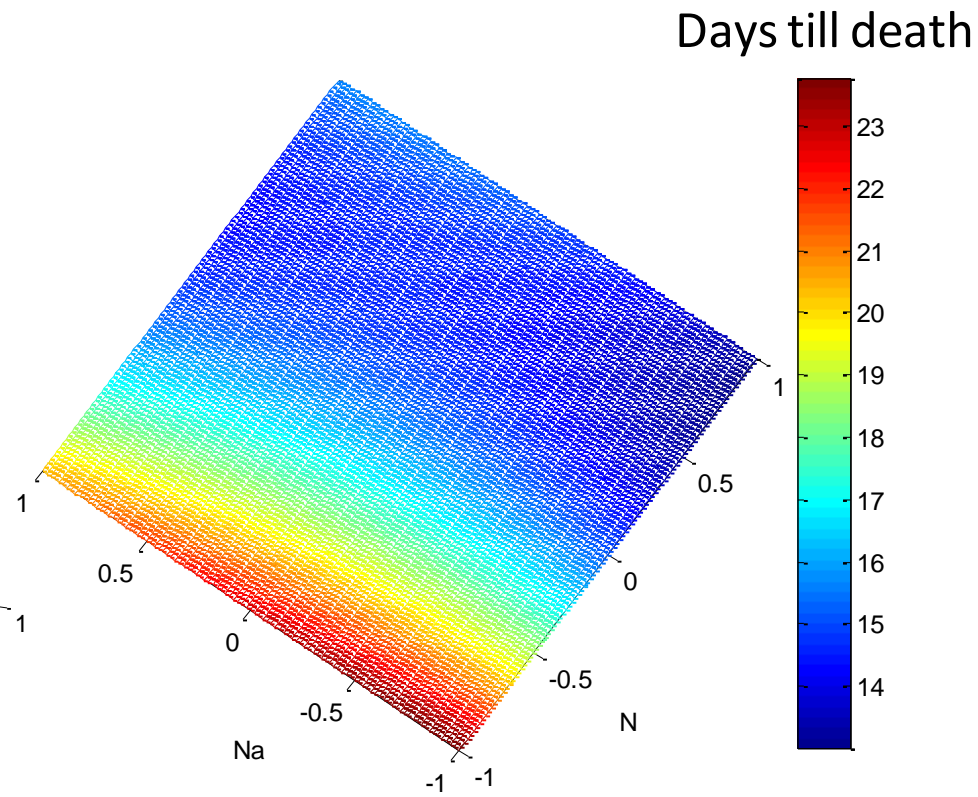
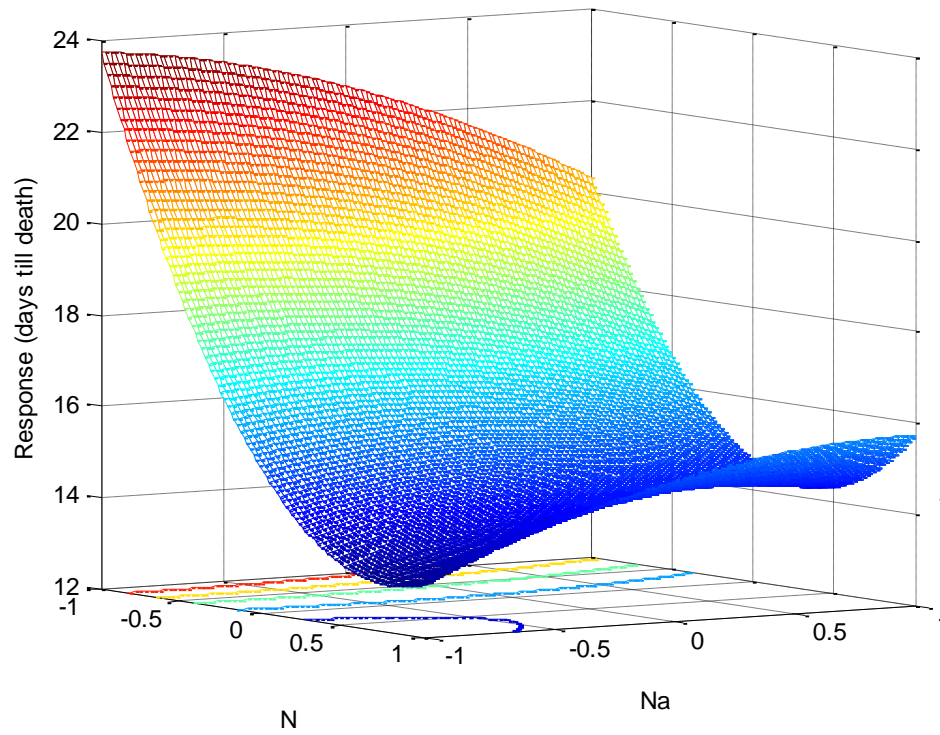


• Design of experiments (DOEs)

To design experiments that can result in the acquisition of good data. Therefore, the relevant information can be extracted on the chemical system under the investigation.

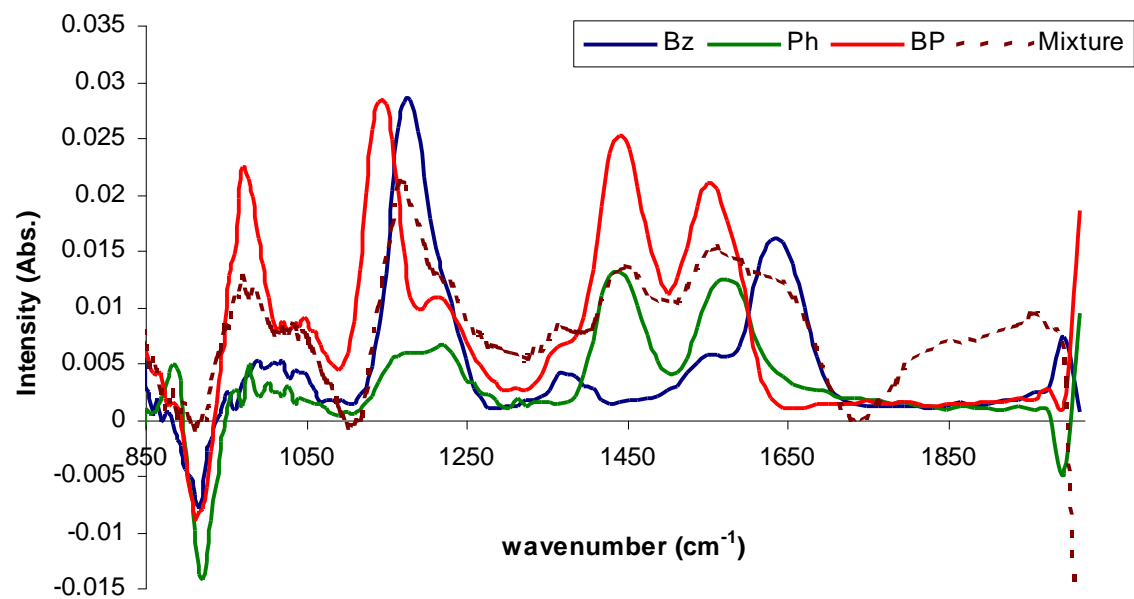
Exp no	Level 1	Level 2
1	+0.71	+0.71
2	+0.71	-0.71
3	-0.71	-0.71
4	-0.71	+0.71
5	+1	0
6	-1	0
7	0	+1
8	0	-1
9	0	0





Resolutions


How many compounds are in this spectra?



Other applications

- Signal processing
- Image analysis
- Feature selection
- Structure-property relationship
- Multiway analysis
- etc.

EARN UP TO \$1000 CASH REBATES FOR FREELANCING
 Sign up for free



'Was a mistake, I'm sorry': Zuckerberg on Cambridge Analytica data breach to CNN

Facebook CEO Zuckerberg said the social media giant would do a full forensic audit if it detected any suspicious activity from here on.

TNM Staff | Thursday, March 22, 2018 - 18:42

Follow @thenewsminute



SOURCE: CNN



BREAKING NEWS
CNN EXCLUSIVE: FACEBOOK CEO MARK ZUCKERBERG BREAKS SILENCE ON "BREACH OF TRUST" BETWEEN SOCIAL NETWORK AND USERS

LIVE
CNN
 6:04 PM PT

AGAINST TRUMP CAN GO FORWARD ▶ YESTERDAY JUDGE ON NY SUPREM AC360°

Bloomberg **MARK ZUCKERBERG**
 FACEBOOK CEO

